■ DIMENSION REDUCTION - Cont'd.

• PCA-based expansion, error etc.

→ The normalized mean squared error is $E_R^{norm} = E_R / \Lambda$.

→ Two (equivalent) conditions can be used to determine whether an R-term approximation is acceptable:

$$\Lambda_R^{norm} > \varepsilon \quad \text{or} \quad E_R^{norm} < 1 - \varepsilon$$

(For data classification, proper values for ε must be determined experimentally. Generally, it is possible that small-value eigenvalues λ_i are relevant for classification.)

• Note: It is appropriate to use the term "information" when referring to eigenvalues and their associated eigenvectors' importance: Shannon's entropy is a measure for information, defined as $H = -\sum_{i=1}^{D} p_i \ln(p_i)$, with normalized probabilities, i.e., $\sum_{i=1}^{D} p_i = 1$. Using the normalized eigenvalues $\lambda_i^{norm} = \lambda_i / \Lambda$ as probabilities p_i, it follows that the eigenvectors $\mathbb{b}_1, \ldots, \mathbb{b}_D$ are optimal as they minimize H. Thus, eigenvector $\mathbb{b}_1$ is most important and $\mathbb{b}_D$ is least important.

• Important issues to be addressed in data classification:

(i) Given a feature point/vector with MISSING data, can one estimate them?

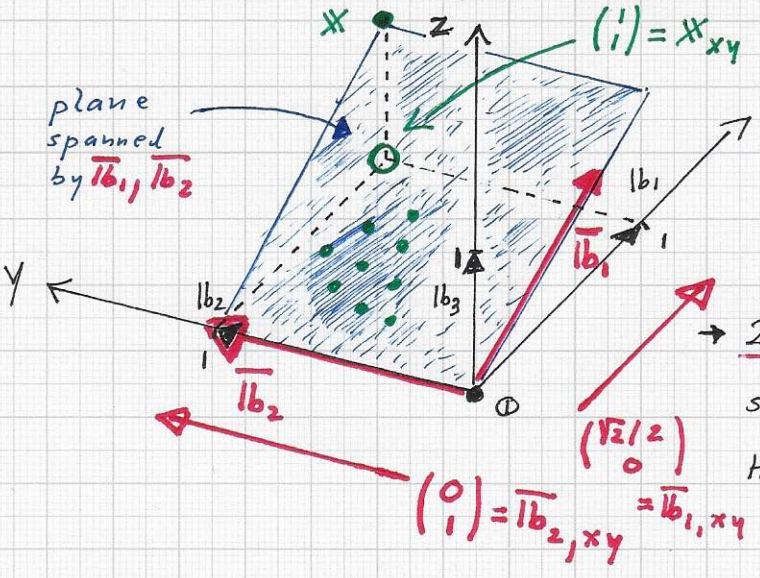
(ii) Given an 'incomplete sample data set,' can one estimate an optimal basis?

(iii) When PCA-based optimal basis construction is not desirable, one should use GRAM-SCHMIDT ortonormalization for basis construction.

DIMENSION REDUCTION – Cont'd.

- **Missing data I**: Knowing a complete PCA basis $\{\mathbb{b}_i\}_{i=1}^{D}$ of a sample data set, how can one approximate missing data values of a sample?

  - **Ex.**: Original 3D data samples lying on a (2D) plane; "reconstruct"/approximate a missing coordinate value



→ $\mathbb{b}_1 = \begin{pmatrix}1\\0\\0\end{pmatrix}$, $\mathbb{b}_2 = \begin{pmatrix}0\\1\\0\end{pmatrix}$, $\mathbb{b}_3 = \begin{pmatrix}0\\0\\1\end{pmatrix}$

→ sample data used to define PCA basis $\{\mathbb{b}_i\}$ lie in a plane

→ 2 PCA basis vectors, $\mathbb{b}_1$ & $\mathbb{b}_2$, suffice to span the plane. Here: $\mathbb{b}_1 = \begin{pmatrix}\sqrt{2}/2\\0\\\sqrt{2}/2\end{pmatrix}$, $\mathbb{b}_2 = \begin{pmatrix}0\\1\\0\end{pmatrix}$

→ The new sample with one missing coordinate value, $\mathbb{x} = \begin{pmatrix}1\\1\\?\end{pmatrix}$, must be reconstructed, defining a "best-possible" reconstruction of $\mathbb{x} = \binom{1}{1}$, lying in the plane.

⇒ Method: Perform BEST APPROXIMATION, in the least-squares sense, to optimally estimate the missing value ?, using $\mathbb{b}_1$ and $\mathbb{b}_2$ as basis vectors for the computed best approximation and "NOT CONSIDERING TERMS OF INNER PRODUCTS BELONGING TO THE ? COORDINATE.

Here: BEST APPROX. OF $\mathbb{x}_{xy}$ WITH $\mathbb{b}_{1,xy}$ AND $\mathbb{b}_{2,xy}$!

■ DIMENSION REDUCTION – Cont'd.

• Missing data I: Using best approximation...

$$\begin{pmatrix} \langle \bar{\mathbf{b}}_1, \bar{\mathbf{b}}_1 \rangle_\bullet & \langle \bar{\mathbf{b}}_1, \bar{\mathbf{b}}_2 \rangle_\bullet \\ \langle \bar{\mathbf{b}}_2, \bar{\mathbf{b}}_1 \rangle_\bullet & \langle \bar{\mathbf{b}}_2, \bar{\mathbf{b}}_2 \rangle_\bullet \end{pmatrix} \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix} = \begin{pmatrix} \langle \underline{\mathbf{x}}, \bar{\mathbf{b}}_1 \rangle_\bullet \\ \langle \underline{\mathbf{x}}, \bar{\mathbf{b}}_2 \rangle_\bullet \end{pmatrix}$$

$$\Leftrightarrow \begin{pmatrix} \begin{pmatrix} \sqrt{2}/2 \\ 0 \\ \bullet \end{pmatrix} \begin{pmatrix} \sqrt{2}/2 \\ 0 \\ \bullet \end{pmatrix} & \begin{pmatrix} \sqrt{2}/2 \\ 0 \\ \bullet \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ \bullet \end{pmatrix} \\ \begin{pmatrix} 0 \\ 1 \\ \bullet \end{pmatrix} \begin{pmatrix} \sqrt{2}/2 \\ 0 \\ \bullet \end{pmatrix} & \begin{pmatrix} 0 \\ 1 \\ \bullet \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ \bullet \end{pmatrix} \end{pmatrix} \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} 1 \\ 1 \\ ? \end{pmatrix} \begin{pmatrix} \sqrt{2}/2 \\ 0 \\ \bullet \end{pmatrix} \\ \begin{pmatrix} 1 \\ 1 \\ ? \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ \bullet \end{pmatrix} \end{pmatrix}$$

$$\Leftrightarrow \begin{pmatrix} 1/2 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix} = \begin{pmatrix} \sqrt{2}/2 \\ 1 \end{pmatrix}$$

$$\Rightarrow \underline{\underline{\begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix} = \begin{pmatrix} \sqrt{2} \\ 1 \end{pmatrix}}}$$

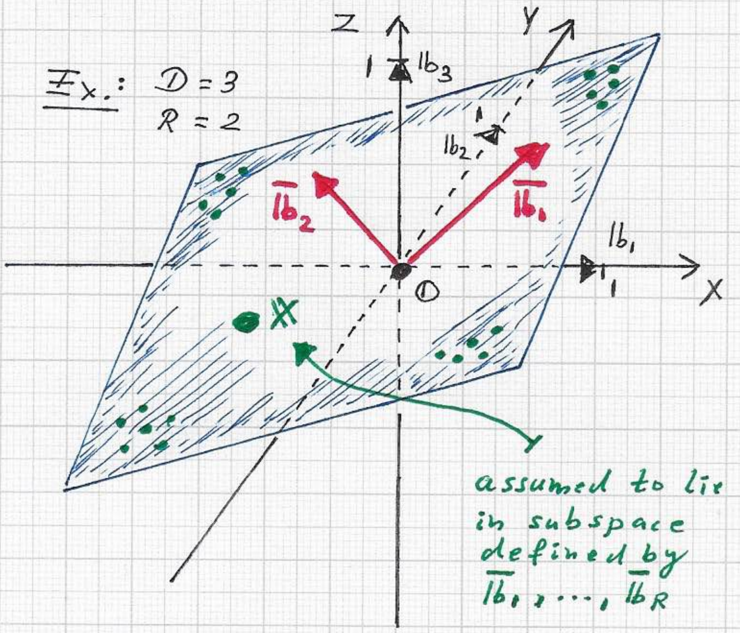
'?' indicates missing value of a coordinate;
'•' indicates that a coordinate is not considered when computing certain inner product terms.

$$\Rightarrow \underline{\underline{\mathbf{x} = \text{"}\mathbf{x}_{app}\text{"} = \sqrt{2} \begin{pmatrix} \sqrt{2}/2 \\ 0 \\ \sqrt{2}/2 \end{pmatrix} + 1 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \text{"}\mathbf{x}_{exact}\text{"}}}$$

• **Notes:**
- Since the datum $\underline{\mathbf{x}}$ is assumed to "behave" like all other sample data $\mathbf{x}$ used to define the PCA basis, this example uses a datum $\underline{\mathbf{x}}$ in the plane of all other sample data.
- Perfect reconstruction ($\mathbf{x}_{app} = \mathbf{x}_{exact}$) is possible in this example, since 2 coordinate values of $\underline{\mathbf{x}}$ are known and $2 \geq$ dimension (plane).

■ DIMENSION REDUCTION – Cont'd.

- Missing data I: Using best approximation...



General case:

→ original data samples $\mathbb{X}$ given in D-dim. space can be represented in the lower-dimensional (R-dim.) space.

→ R-dim. space is spanned by basis vectors $\overline{\mathbb{b}}_i$, $i=1 \ldots R$, resulting from PCA.

→ new sample $\underline{\mathbb{X}}$ has one or multiple missing coord. values.

→ Use best approximation to estimate (sometimes reconstruct) the missing values of $\underline{\mathbb{X}}$, resulting in the approximation $\mathbb{X}_{app}$ of $\mathbb{X}$.

→ Compute the best approximation $\mathbb{X}_{app}$ by considering only the "valid" coordinates for all inner product terms, i.e., do not consider coordinates that are missing in $\underline{\mathbb{X}}$. Use $\overline{\mathbb{b}}_1, \ldots, \overline{\mathbb{b}}_R$ as basis vectors for the best approximation computations. In this basis, the approximation $\sum_{i=1}^{R} \overline{x}_i \overline{\mathbb{b}}_i$ results, having coord. values for all D original coordinates. $\underline{\mathbb{X}}$ has missing coord. values (? values); the missing values are defined by the respective $\overline{x}_i$ coefficients in the expansion $\sum_{i=1}^{R} \overline{x}_i \overline{\mathbb{b}}_i$.

■ DIMENSION REDUCTION – Cont'd.

- Missing data I: best approximation...

→ Compute R coefficients by solving the system

$$\begin{bmatrix} \langle \bar{\mathbb{b}}_1, \bar{\mathbb{b}}_1 \rangle_\bullet & \cdots & \langle \bar{\mathbb{b}}_1, \bar{\mathbb{b}}_R \rangle_\bullet \\ \vdots & & \vdots \\ \langle \bar{\mathbb{b}}_R, \bar{\mathbb{b}}_1 \rangle_\bullet & & \langle \bar{\mathbb{b}}_R, \bar{\mathbb{b}}_R \rangle_\bullet \end{bmatrix} \begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_R \end{bmatrix} = \begin{bmatrix} \langle \underline{\mathbb{x}}, \bar{\mathbb{b}}_1 \rangle_\bullet \\ \vdots \\ \langle \underline{\mathbb{x}}, \bar{\mathbb{b}}_R \rangle_\bullet \end{bmatrix}$$

generally not a diagonal matrix!

- $\langle\ \rangle_\bullet$ does NOT consider inner product terms that correspond to missing "? values" of $\underline{\mathbb{x}}$.
- Consequently, even though $\{\bar{\mathbb{b}}_i\}_{i=1}^{R}$ is an orthonormal basis the values of all $\langle\ \rangle_\bullet$ products can be different from zero!

→ The resulting approximation $\sum_{i=1}^{R} \bar{x}_i \bar{\mathbb{b}}_i$ produces the desired D coordinate values.

→ **The best approximation $\mathbb{x}_{app}$ of $\mathbb{x}$ is that approximation using all the coordinate values known for $\underline{\mathbb{x}}$ and uses those coefficients $\bar{x}_i$ for the missing "? values" of $\underline{\mathbb{x}}$.**

(Note: The dimensions of the original space, D, and the PCA-defined space, R, and the number of of missing values, "? values," of $\underline{\mathbb{x}}$ determine the approximation error of $\mathbb{x}_{app}$. Ofen, $\mathbb{x}_{app}$ is a perfect reconstruction of $\mathbb{x}$.)

BH